达尔文是干什么的，以及语言、教育和智力研究

作者：顾立帆

亚里斯多德是生物分类学家，孟德尔是遗传学家。达尔文是干什么的？准确一点说是演化学家。演化学主要是生物历史。也就是说达尔文主要是生物历史学家。

但演化学不仅仅是生物历史研究，也是生物科学研究。新冠病毒流行后，演化树和证据链成了病毒研究的关注焦点。以后太空宇航大规模发展起来后，在高辐射、重力多变等等更复杂的环境里，演化会更加快速和复杂。也就是说，达尔文的演化理论也会越来越重要。

亚里斯多德生物分类法以外在形态为基础，有很多误导。现代生物分类法以达尔文的演化理论为基础，按基因演化历史分类，但有新问题。选择哪些演化分支点作为分类基础非常主观，出现很多不同的分类方法。唯一有严格规范的物种定义要求不同物种间有生殖隔离。而在基因技术突飞猛进的时代，生殖隔离要求恰恰会较容易打破，导致更大的混乱。

生物分类法本来希望成为一种标准，但实际上是语言的选择，结果出现语言的多样化。现在生物分类五花八门，分类方法越来越多。这跟化学元素周期表和量子力学标准模型的定义严谨、层次分明形成明显对照，显示了生命科学跟物理科学有本质区别。物理科学包括经典物理、化学、量子物理等等，生命科学包括生物学、生理医学、心理学等等。

如果生物分类法不得不允许多样化，并非以基因演化为基础的郑作新鸟类分类法是不是还有存在价值？郑作新分类法跟以形态为基础的传统分类法也有些不同，代表一种较典型的语言思维习惯。游禽、涉禽、攀禽、陆禽是较好的生态分类，鸣禽、猛禽却是跟前面四种生态分类不同的典型分类。这种混杂的分类法，也是深度学习的典型特征和问题。

事实上，类似游禽、涉禽、攀禽、陆禽等等的生态模式不是鸟类独有的。生态模式和基因的关系和互动也远比想象的复杂。

生命科学研究中，Molecular Clock 的有效范围也有争议。如果说在物理科学和生命科学中，概率统计方法只是有争议，在智力研究中概率统计方法会产生严重问题则可通过实验验证。缺乏准确的逻辑临界点分析，谈论智力研究正确率有多高没有意义。在模式不稳定演变下，正确率会急剧变化。

深度学习则还有更本质的问题：行为主义（Behaviorism）的缺陷。行为主义本来是心理学的一种学派，强调训练、强化和控制。种植业和制造业的很多工作，大部分人通过训练确实可以做，这是行为主义的基础。但科学研究和其它很多行业的很多工作，大部分人通过训练仍不适合，只有少数有特长的人才能胜任。而且能胜任某些研究的人不一定适合其他研究，这在科学时代成为行为主义的致命缺陷。

控制是动物生理层次的智力，对人类实施控制会造成其思维的降阶和降维等等问题，在某种程度上损害人类特有智力。所以对人类的控制术实际上是一种 tradeoff。新科学的探索需要更多地发挥自由意志的优势，尽可能避免控制术带来的危害。

基于生命的行为主义失败了，但人工智能又给行为主义带来新的狂热：训练、强化和控制真能造就万能的人工智能？有人工智能专家说“人类正在扮演上帝的角色”。Demis Hassabis 则宣称 Deepmind 的技术是 potentially a meta-solution to any problem。Deepmind 最近的观点是 Reward is Enough。这些在学术上并不可能。

“扮演上帝的角色”不但需要准确理解人类智力、准确解释 causation 关系，还需要准确理解上帝的逻辑。人工智能专家能做到吗？Demis Hassabis 真能解决 any problem，包括裁决信仰争议，解决波函数坍塌、多世界诠释等等量子力学诠释争议和其它各种科学争议？

如果 Demis Hassabis 的技术真能解决 any problem，用这个技术不是就能建立更优越的普世计划经济模式？事实上就有一种从垄断资本主义到社会主义的理论。计划经济恰恰就是计算思维的产物，罗马俱乐部的人口理论也是（笔者不反对计划经济，只是认为计划经济不是普世模式）。人工智能专家没有意识到这一点，是因为缺乏良好的通识教育。

如果 AlphaCode 真能完成人们正在编写的任何内容，Alphabet/Google 可以削减其它部门的程序员，只留 Deepmind 研究人工智能；世界也不需要很多程序员和其他工作者，大学未来只需培养去 Deepmind 工作的人工智能算法研究者；硅谷和美国都不需要那么多就业者，可以

朝底特律方向发展。

Demis Hassabis 和 Deepmind 不懂什么是人工智能无法做的，却误导科学。大量研发资金投向前景不大的方向，而纠正严重学术错误、探索新科学的努力却被封杀，这是教育的失败和学术界的耻辱。公众受愚弄，高房价、高收入还能维持下去吗?

由此可见他们的语言智力和逻辑判断能力理应受质疑。这也暴露了计算思维和一阶逻辑训练的局限性。事实上，他们的语言处理技术缺乏抽象思维、语义演变、高阶逻辑、诡辩分析和判断力等等方面的有效测试。他们的技术连较低目标：准确理解人类智力都做不到。

智力现象比生命演化现象节奏快得多，问题也更严重。自动驾驶标准中的问题会导致的后果当然比生物分类法严重得多。现有标准没有找到关键的有效观察量，也无法有效测试。在模式不稳定演变下会出现严重问题。根据自由意志的研究就可以证明 SAE level 4 自动驾驶不可能实现，当然还有更严重的问题。

事实上，人类的教育模式、经济模式及思维和智力模式的不稳定发展和演变远比生态模式和基因的互动更加复杂。现有学术方法无法处理这种不稳定性，需要新的理论和方法来研究逻辑临界点的演变和发展以便处理这种不稳定演变。目前缺乏对 AlphaGo Zero 和各种后续系列以及自动驾驶汽车的逻辑临界点分析和实验。

哥德尔批评过逻辑实证主义（logical positivism 或 logical empiricism）的问题，这跟学术语言的使用规范有关。不正确的语言使用方法会严重误导科学实验的设计和对实验结果的解释，误导科学发展。这些在智力相关研究中正变得越来越严重，并危及科学和文明基础。

笔者研究的是智力科学，主要是人类智力的发展。跟生物学不同，人类智力发展进程中，基因变化较小，主要是语言、数学、哲学、科学等等的发展。

计算机是数学的一个子集。增强某些方面的数学功能本来就是研发计算机的主要目的。但数学不能覆盖一些重要的语言和科学功能，计算机也不能覆盖一些重要的数学功能。

Universal Approximation Theorem 和 Nash Embedding Theorems 适用条件有限制，各种计算机语言模型包括 word-embedding Vector Space 无法处理抽象思维、语义演变、高阶逻辑、诡辩分析和判断力等等问题。这是错误的关键。

智力科学需要建立在严格控制条件的实验基础上。笔者初步设计的实验，主要是人工智能的围棋实验和语言实验。以后有条件的话逐步扩展。

Deepmind 对 AlphaGo Zero 及各种后续系列参与严谨科学实验的公开性很低，所以笔者也只能有条件地逐步公开实验细节，以保证人工智能实验的有效性，以免实验受人类智力污染。

目前测试语言处理的各种数据集，缺乏很多对科学和文明发展非常重要和关键的抽象思维、语义演变、高阶逻辑和诡辩分析等等能力的测试，明显高估人工智能而有很严重的误导。

笔者整理了一些相关的语言测试案例。因为研究工作得不到报酬、计算机长年被入侵和监控、也没有安全的居住环境，所以笔者不能把这些语言测试案例在计算机或纸面上记录下来，以免工作成果受人窃取而受奴役。笔者记忆力也不强，时间一长，很多设计好的语言测试案例已经遗忘了。随着笔者健康的日益衰退，以后能回忆起多少也不知道。

所以，恶性竞争和奴役会严重阻碍科学和文明的发展。

按媒体报道，目前无人汽车 SAE level 4 的测试实际上是按地区测试并按地区宣布通过测试的，也就是实际上在测试中分成一些典型的类型如：亚利桑那人口密度较低、气候晴朗干燥地区，旧金山人口密度较高、雨季多雨地区，以及更复杂的风雪和沙尘等等。这种测试上的分类方法有什么问题，忽略了什么关键因素?

模拟测试无法真正模拟人类智力，特别是其不稳定模式演变。在实用场景下，跟无人汽车相关的人类智力演变相对较慢，模式不稳定演变会发生在大规模应用后较长时间、普通人对无人汽车特点和缺陷了解比较透彻之后。太迟发现问题会造成严重后果。所以需要研究智力科学的基础理论和实验。

笔者按学术自由发表学术观点。别人想准确了解，应按合法途径、经过身份认证向笔者直接确认。捕风捉影、扭曲猜测者，责任自负。按学术界惯例，对尚未在正式学术论文中发表的想法，笔者甚至无需解释。所以，想封杀笔者的人，也责任自负。

[illegible] leukotomy [illegible]

[illegible] game[illegible]

[illegible]

[illegible] reward[illegible] Deepmind [illegible] Reward is Enough [illegible]

[illegible] A Treatise on Probability [illegible] causation [illegible]

[illegible]

[illegible]causation [illegible]

[illegible]

1[illegible]
2[illegible]
3[illegible]

[illegible]“Confucius taught that marriage lies at the foundation of government.”[illegible] causation [illegible]

[illegible]Marc Aurel Stein[illegible]John Leighton Stuart[illegible]

[illegible] causation [illegible]

[illegible]

[illegible]Demis Hassabis [illegible] Deepmind [illegible]

[illegible]

附录一、语义演变和复杂性分析的一个案例

这是一个在人类文明史上出现过的曾经非常重要的案例。以此为 context 可以设计一些测试问题来区分人类和人工智能以及检测人工智能的某些语言处理能力。

《管子》分析了皇道、帝道、王道、霸道四种模式。嬴政试图融合皇道和帝道，创立皇帝名位。自此皇和帝的语义已有新发展，不同于先秦的皇道和帝道。

项羽、韩信试图恢复王道。项羽先前的诸侯上将军接近霸道，后来的霸王接近王道中的天子。韩信想做的王其实是王道中的诸侯。但两人都失败。刘邦则在皇帝模式和王模式之间妥协，保留同姓诸侯。

汉宣帝说“霸王道杂之”和《管子》中“明一者皇，察道者帝，通德者王，谋得兵胜者霸。”的王、霸语义既有关联又明显不同。前者指皇帝模式下两种思维、运作方式，包括皇帝驾驭臣下的方法。后者指不同于皇帝模式的两种模式。

皇帝模式下通德者当然不能自行称王。后来王莽则按《管子》语义“通德者王”篡位。霸的语义演变更复杂。《孟子》解释：“霸者，长也。言为诸侯之长。”《白虎通》进一步说明其语义演变的来源：“霸者，伯也，行方伯之职。”伯是家族中长者。霸跟伯同义，引申为诸侯中长者的意思。《管子》说诸侯靠谋得兵胜称霸，这当然不是汉宣帝所说霸的语义。准确理解这些语义演变和发展，才能真正理解中国历史发展的关键。

司马光评论汉宣帝的话说“王霸无异道”。这有两点争议。一是根据上述分析，他混淆了汉宣帝所说和《管子》所说王、霸语义的不同。二是他试图说《管子》中王和霸本质上都是“本仁祖义，任贤使能，赏善罚恶，禁暴诛乱。”还说：“天子微弱不能治诸侯，诸侯有能率其与国同讨不庭以尊王室者，则谓之霸。”这跟事实明显不符。春秋霸主楚庄王都自称王了，岂是“同讨不庭以尊王室”？楚庄王虽自称王，别人仍称他为霸。

所以很多语汇既不能拘泥于原定义、也不可以随意重新定义，而要根据事实分析语义的演变。这些复杂的王和霸的用法和关系，人工智能能理解、分析清楚吗？人工智能是否能找出从《管子》所说到汉宣帝所说王、霸语义演变的原因及逻辑临界点？

根据此类语义分析案例，笔者可以设计很多语言测试的方法，来测试那些对文明和科学发展非常重要却被现在的语言测试集忽视的问题，以准确测试人工智能的智力。本案例涉及历史知识，更广泛的测试需要各种专题的问题，也需要较少涉及专门知识的问题。

类似抽象了静止和运动的惯性以及哈密顿量等等概念形成了物理科学的基础，智力科学也需要许多新概念来有效设计实验、发现规律。围绕这些概念，也可在很多领域设计问题。

附录二、关于科学验证方法的退化

由于很多物理科学前沿研究问题越来越难以验证，近年科学界有人主张证明的概念也需要演化。但他们提出的新验证方法实际上是退化版的，而不是进化版的，典型的表达如：

> First, if scientists have tried, and failed, to come up with an alternative theory that explains a phenomenon well, that counts as evidence in favor of the original theory. Second, if a theory keeps seeming like a better idea the more you study it, that’s another plus-one. And if a line of thought produced a theory that evidence later supported, chances are it will again.

这种方法并没有依严谨的逻辑去证伪其它可能来真正证明一理论，明显达不到科学验证的标准。自然研究经历了自然历史（Historia Naturalis）、自然哲学（Philosophiae Naturalis）、自然科学（scientia naturalis）的发展。这种退化版的方法仅属于自然历史范畴的研究。

这也是物理科学研究正遭遇越来越严重的瓶颈，科学思维的严谨性正在退化的证据。所以需要智力新科学来研究这种现象，并通过严谨的智力科学实验带动科学严谨性的继续发展。

附录三、对多重宇宙和量子力学诠释争议的智力结构分析和语言处理

[illegible]

[illegible]

[illegible]

[illegible] Are there really many worlds in the "Many-worlds interpretation" of Quantum Mechanics?[illegible]the development of «decoherence theory» revealed that, using the standard formalism of quantum mechanics, macroscopically distinct branches of the wavefunction were almost entirely free from interference and evolve approximately classically[illegible] almost [illegible]

[illegible]The Many-worlds Interpretation[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]…[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible] Selfish gene [illegible]

[illegible]

[illegible]

[illegible]Waymo [illegible] SAE level 4 [illegible]

[illegible]Waymo [illegible] crash data [illegible] trade secret[illegible] data [illegible]

[illegible]Deepmind [illegible] Waymo [illegible]

[illegible] D-wave [illegible] Google Quantum Supremacy [illegible]

[illegible]